

**Inter-rater reliability and a training effect of the functional movement screen in police physical training instructors**

Conkin, Charise; Hinton, Ben; Ross, Kayla; Schram, Ben; Pope, Rodney; Orr, Robin

*Published in:*  
Cogent Social Sciences

*DOI:*  
[10.1080/23311886.2020.1763769](https://doi.org/10.1080/23311886.2020.1763769)

*Licence:*  
CC BY

[Link to output in Bond University research repository.](#)

*Recommended citation(APA):*  
Conkin, C., Hinton, B., Ross, K., Schram, B., Pope, R., & Orr, R. (2020). Inter-rater reliability and a training effect of the functional movement screen in police physical training instructors. *Cogent Social Sciences*, 6(1), [1763769]. <https://doi.org/10.1080/23311886.2020.1763769>

**General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

For more information, or if you believe that this document breaches copyright, please contact the Bond University research repository coordinator.



Received: 05 August 2019  
Accepted: 28 April 2020

\*Corresponding author: Robin Orr,  
Faculty of Health Science & Medicine,  
Bond University, Gold Coast, QLD  
4229, Australia  
Email: [rorr@bond.edu.au](mailto:rorr@bond.edu.au)

Reviewing editor:  
Heng Choon (Oliver) Chan, City  
University of Hong Kong, Hong Kong

Additional information is available at  
the end of the article

## LAW, CRIMINOLOGY & CRIMINAL JUSTICE | RESEARCH ARTICLE

# Inter-rater reliability and a training effect of the functional movement screen in police physical training instructors

Charise Conkin<sup>1</sup>, Ben Hinton<sup>2</sup>, Kayla Ross<sup>2</sup>, Ben Schram<sup>1,3</sup>, Rodney Pope<sup>3,4</sup> and Robin Orr<sup>1,3\*</sup>

**Abstract:** The aim of this study was to determine the inter-rater reliability of the Functional Movement Screen (FMS<sup>TM</sup>) within a police population and determine whether formal training improved reliability and assessment accuracy. Police Physical Training Instructors (PTI) (n = 67) rated 98 randomised videos of officers performing four primary FMS<sup>TM</sup> tasks (overhead squat, hurdle step, in-line lunge, and rotary stability) twice within a two-day annual training program. A one-hour FMS<sup>TM</sup> training session was completed between the two assessment periods. PTI scores were compared to a Master score. The level of agreement between raters was low to moderate across all FMS<sup>TM</sup> items. The inter-rater reliability between the raters did not improve significantly following training. The level of agreement between “Rater” and “Master” scores improved significantly post-training. FMS<sup>TM</sup> subtest items have varying reliability. Staff training should be performed and compared to a Master rater prior to employing the FMS<sup>TM</sup> as an assessment tool in law enforcement.

**Subjects:** Law; Sports and Leisure; Health and Social Care

**Keywords:** reliability; movement screen; physical training instructor; law enforcement

### ABOUT THE AUTHOR

Associate Professor Rob Orr served in the Australian Army for over 23 years as an infantry soldier, Physical Training Instructor, physiotherapist and human performance officer. Still serving in the Army Reserve, Dr Orr is the director of the Tactical Research Unit at Bond University, teaches in the Doctor of Physiotherapy Program, and supervises masters and doctoral students. Rob has run training for military, law enforcement, and firefighter agencies across the globe providing research, consultancy and educational services. With awards for research outcomes, over 100 tactical publications and numerous annual speaking conference and workshop requirements, Dr Orr serves as the editor of the Tactical Strength and Conditioning Report and on several scientific and organising committees for renowned international conferences.

### PUBLIC INTEREST STATEMENT

Police services, and other similar occupations use the Functional Movement Screen (FMS<sup>TM</sup>) to assess physical movement competency. With the importance that this screening tool plays in a police officer's service, ensuring different assessors would score an officer the same (inter-rater reliability), is of importance. So too is determining whether training in this tool is needed to improve reliability. The study found that the FMS<sup>TM</sup> had low to moderate reliability, which does not improve with FMS<sup>TM</sup> training. However, the individual scores of assessors were closer to the master (correct) score following training. These results suggest that FMS<sup>TM</sup> training may not improve inter-rater reliability but may improve accuracy in scores. As such, when staff are being trained to implement the FMS<sup>TM</sup> their scores should be compared to a “master” score to improve score quality.

## 1. Introduction

The nature of law enforcement sees police officers perform tasks that can be unpredictable and in potentially dangerous or hostile environments (Bonneau & Brown, 1995). Tasks can include “attending a domestic incident” to “affecting an arrest” (Decker et al., 2016) and require officers to pursue suspects on foot to prevent escape and/or the use of reasonable force such as wrestling with an offender (Achterstraat, 2008). The physical nature of these tasks leaves officers at risk of musculoskeletal injury (Peate et al., 2007). As such, police officers are known to suffer a range of musculoskeletal injuries, with soft-tissue sprains and strains commonly caused by a non-compliant offender serving as an example (Lyons et al., 2017).

The Functional Movement Screen (FMS)<sup>TM</sup> is an assessment tool used in screening for musculoskeletal injury risk in both sporting (Chorba et al., 2010; Kiesel et al., 2011; Letafatkar et al., 2014) and tactical populations (Bushman et al., 2015; Lisman et al., 2013; Peate et al., 2007) and has been associated with some elements of occupational performance in law enforcement (Bock, Stierli, Hinton & Orr, 2016). This tool measures the musculoskeletal movement ability of a participant to identify functional limitations and asymmetries through the use of seven key movement patterns: a deep squat, hurdle step, in-line lunge, push-up and rotary stability task (Cook et al., 2006). The FMS<sup>TM</sup> is designed to place the body in challenging positions that may reveal weaknesses, imbalances and compensations should the participant not have sufficient stability (Cook et al., 2006). The FMS<sup>TM</sup> also includes three clearing tests including the shoulder mobility, extension and flexion-clearing tests, which are graded either positive or negative based on pain. On this basis, the highest score than can be attained in the FMS<sup>TM</sup> is 21 points.

Previous research has suggested that lower scores, being <14 points, are associated with an increased risk of injury (Chorba et al., 2010; Kiesel et al., 2007; Perry & Koehle, 2013; Schneiders et al., 2011). Two reports (Chorba et al., 2010; Kiesel et al., 2007) found that athletes who scored <14 points had a significantly greater risk of injury when participating in sports. In addition, reports within general populations (Perry & Koehle, 2013; Schneiders et al., 2011) have confirmed a FMS<sup>TM</sup> score of <14 points is indicative of an increased risk of injury. In tactical populations, marine officer candidates who scored poorly in the FMS<sup>TM</sup> (<14 points) were more likely to experience an injury during Marine Corps officer training (Lisman et al., 2013; O'Connor et al., 2011).

Considering this, if the FMS<sup>TM</sup> is to be used as a predictor of injury in a law enforcement population, inter-tester reliability is paramount, especially if departmental decisions are going to be based on assessment outcomes. The FMS<sup>TM</sup> was constructed so that individuals of any skill level could be trained to correctly implement the tool. So far, research has shown that physical therapists and physical therapy students, athletic trainers and athletic trainer students, Certified Strength and Conditioning Specialists and FMS<sup>TM</sup> certified specialists have good to excellent reliability in scoring the FMS<sup>TM</sup> (Butler et al., 2012; Minick et al., 2010; Onate et al., 2012; Smith et al., 2013; Teyhen et al., 2012). This research has been conducted on middle school-aged students (Butler et al., 2012; Wright et al., 2015), youth hockey players (Parenteau et al., 2014), physically active adults (Onate et al., 2012; Schneiders et al., 2011), active-duty service members (Teyhen et al., 2012), elite squash players (Leeder et al., 2016), NCAA Division I varsity athletes (Shultz et al., 2013), college students (Gulgin & Hoogenboom, 2014; Minick et al., 2010; Sorenson, 2016), university students (Gribble et al., 2013), and physical therapy students (Smith et al., 2013), with the number of raters ranging from two (Butler et al., 2012; Onate et al., 2012; Schneiders et al., 2011) to 38 (Gribble et al., 2013) and participants from three (Gribble et al., 2013) to 209 (Schneiders et al., 2011).

As yet, the inter-rater reliability of the FMS<sup>TM</sup>, conducted by Physical Training Instructors (PTI) serving within a law enforcement agency has not been determined. The importance of determining the level of agreement between PTIs is of note given that different PTIs, from the same or different locations, could assess officers and their findings may impact on the officer and departmental resources. On this basis, the aim of this study was to determine the inter-rater-reliability of the FMS<sup>TM</sup> within a law enforcement PTI population and to determine whether providing formal training in the FMS<sup>TM</sup> improved inter-rater reliability and accuracy.

## 2. Materials and methods

In this retrospective cohort design, a total of 14 (6 women, 8 men) police officers were examined. Raters were recruited from an Australian state police force whose officers were attending a Physical Training Instructor Annual Training session. Of the 67 police PTI training attendees, all 67 consented to participate in the research (22 women: age =  $39.3 \pm 6.1$  years; 45 men: age =  $39.3 \pm 7.9$  years). Criteria for participants were: 1) participant must be a serving police employee, and 2) be a qualified PTI. Participant data were excluded if: a) the participant's script could not be clearly determined, or b) the participant did not score according to the FMS<sup>TM</sup> scoring criteria (e.g. gave  $\frac{1}{2}$  scores). Level of experience as a PTI ranged from 1 to 9 years and all PTI's had some level of experience with the FMS<sup>TM</sup>. The sample of raters was representative of the population. All participants provided written informed consent, and ethics approval for this study was provided by the Bond University Human Research Ethics Committee under protocol RO 1595. Gatekeeper approvals were provided by the Australian state police force from which participants were drawn.

Four of the seven FMS<sup>TM</sup> tasks were previously recorded displaying a frontal and sagittal view using two Apple iPhone 6 mobile phones set up on a tripod. The position of participants and camera set up was standardized for each view. The videos showed a split view of 14 unidentified police officers performing the overhead squat, in-line lunge, hurdle step and rotary stability with each exposure played three times, followed by 10 seconds in which to score. The push-up assessment, shoulder mobility and active straight leg raise were excluded from this study as the scoring methods needed direct interface with the participant and as such could not be scored by video. Each exposure was presented in a predetermined randomized order and changed from pre- to post-testing. Participants were given a score sheet in which to record their scores. The movement patterns were scored on a scale of one (1) to three (3) based on the quality of movement. A score of zero (0) was unable to be used, as pain is unable to be determined during video data capture.

Over 2 days, the raters were required to attend two FMS<sup>TM</sup> rating sessions (Session 1 and 2) and a FMS<sup>TM</sup> education session interspersed between the rating sessions. Session 1 was held on the first day of the conference and consisted of a 50-min data capturing session where the participants scored a total of 98 videos of a randomized sequence of unknown police officers performing the FMS<sup>TM</sup>.

Following this first session, in controllable groups of approximately 20 participants underwent a FMS<sup>TM</sup> training session. Education on the procedure and scoring of the FMS<sup>TM</sup> was provided by the same two researchers (BH and KR) who determined a "Master" score. Both researchers are Physical Training Instructors who have over 5 years' of PTI experience, a Level 1 and Level 2 FMS<sup>TM</sup> Certification, and have each conducted over 2,000 Functional Movement Screens in this population. Each group was provided with the same one-hour lecture on the FMS<sup>TM</sup> movements, scoring, and variable differences. After the education session, the raters were again exposed to a different randomized sequence of the 98 videos and same process as Session 1.

For both "pre-training" and "post training" descriptive analysis and inter-rater reliability, levels of agreement were determined across all raters for each test element, including R, L, and Final Scores where relevant ( $n = 14$  videos for each element). Due to the ordinal nature of individual FMS<sup>TM</sup> item scores, Kendall's Coefficient of Concordance ( $W$ ) and the Average Spearman Correlation Over All Raters ( $\bar{R}_s$ ) were calculated to determine the levels of agreement across all raters in the scores they assigned for each element of each assessed FMS<sup>TM</sup> item, where each element of each item was represented by and assessed across 14 different video clips.

Average Spearman Correlations were then calculated to determine levels of concordance between "Master" scores and rater scores for each FMS<sup>TM</sup> item (with multiple videos and associated scores for each item from each individual rater) both pre- and post-training. The "Master" score was determined by the two researchers (BH and KR) as mentioned above. Frequency distributions (charts and tables) of numeric differences between rater scores and master scores pre- and post-training were compiled to visually determine the extent to which rater scores

differed from master scores (as a measure criterion-related validity), and to determine if these distributions changed with training. Related Samples Wilcoxon Signed Rank Tests were then conducted to examine the statistical significance of any visually apparent differences in these frequency distributions.

### 3. Results

In total, 67 raters scored 98 different video clips for each FMS™ item (Overhead Squat,  $n = 14$ : Hurdle Step Left,  $n = 14$ , and Right,  $n = 14$ : In line Lunge Left,  $n = 14$ , and Right  $n = 14$ , Rotary Stability Left,  $n = 14$ , and Right,  $n = 14$ ). The resulting Kendall's Coefficients of Concordance ( $W$ ) and Average Spearman Correlations ( $\bar{R}_s$ ) overall raters pre- and post-training are shown in Table 1. These results suggest that the level of agreement between raters was low to moderate across all items with rotary stability on the right side having the lowest reliability pre-training and post-training (Table 1). The left inline lunge and hurdle step demonstrated the highest reliability across the FMS™ items pre- and post FMS™ training, respectively. Considering this, it can be seen that the levels of agreement across all raters improved post-training in 50% of FMS™ sub-items (overhead squat, left hurdle step, left, right and final rotary stability), but were poorer post-training in the other 50% (right hurdle step, final hurdle step, left, right and final inline lunge). The difference in levels of agreement of FMS™ sub-items pre-training to post-training were largest in the overhead squat, inline lunge on the left, inline lunge total, rotary stability on the right and final rotary stability scores. Overall, the hurdle step had the smallest changes in pre-training to post-training FMS™ sub-item scores.

While the inter-rater reliability between the general "Raters" did not improve consistently following training (indicating some ongoing variability in scores allocated for specific video clips), the level of agreement between "Rater" and "Master" scores did significantly improve post-training for each of the FMS™ items assessed (Table 2). This suggests that, following training, "Raters" allocated scores that were much more consistent with Master scores for the same FMS™ items.

### 4. Discussion

The aim of this study was to determine the inter-rater reliability of the FMS™ within a law enforcement PTI population and to determine whether providing formal training in the FMS™ improved inter-rater reliability and accuracy. The results suggest that the inter-rater reliability between the four individual components of the FMS™ assessed were only low to moderate, improving to moderate, following training. While not significantly improving rater scores post-training, the variability when comparing pre-training and post-training to the "Master" score did

**Table 1. Inter-rater reliability of FMS™ subtest items pre- and post- FMS™ training**

	Pre-training		Post-training	
	Kendall's $W$	Average spearman correlation	Kendall's $W$	Average spearman correlation
Overhead squat—Final	0.323	0.313	0.511	0.504
Hurdle step—Left	0.573	0.567	0.585	0.579
Hurdle step—Right	0.56	0.553	0.535	0.528
Hurdle step—Final	0.583	0.577	0.548	0.553
Inline lunge—Left	0.642	0.637	0.455	0.446
Inline lunge—Right	0.432	0.424	0.426	0.417
Inline lunge—Total	0.512	0.505	0.404	0.395
Rotary stability—left	0.514	0.507	0.548	0.541
Rotary stability—Right	0.223	0.211	0.365	0.355
Rotary stability—Final	0.374	0.365	0.504	0.496

**Table 2. Inter-rater reliability between general “raters” and “master” score**

	Average spearman correlation			
	Between rater and master scores		% Rater scores matching master scores	
	Pre-training	Post-training	Pre-training	Post-training
Overhead squat	0.392	0.462	34%	59%**
Hurdle step	0.461	0.584	42%	77%**
Inline lunge	0.417	0.482	29%	64%**
Rotary stability	0.17	0.549	36%	78%**

\*\*p <.001 for difference, assessed by Wilcoxon Signed Rank Test.

improve meaning that the PTI scores were more closely related to the “Master” score. This study was different from others in multiple ways: the use of PTIs as raters, introducing a FMS™ training between sessions, and comparing raters of variable experience to Master raters.

No other known studies have examined the reliability of the FMS™ using police PTIs as examiners. The majority of studies examine the inter-rater reliability of the FMS™ using physical therapists, athletic trainers, Certified Strength and Conditioning Specialists and FMS™ certified specialists. The closest example to law enforcement includes the reliability of physiotherapy students rating the FMS™ on armed service members (Teyhen et al., 2012); however, no service members were used to rate the FMS™. In general, the reliability of the FMS™ demonstrates a high level of inter-rater reliability for the FMS™ composite score and a moderate to a high level of inter-rater reliability of the FMS™ items (Bonazza et al., 2017; Cuchna et al., 2016; Stobierski et al., 2015). A FMS™ composite score was not calculated in this study due to the removal of three FMS™ test items, however, the remaining FMS™ items had an overall low to moderate reliability ( $r = 0.425$ ). The overall FMS™ reliability in this study using four items was lower than documented in previous literature. Composite or FMS™ total reliability scores have found acceptable inter-rater reliability with ICC values of 0.76–0.98 (Bonazza et al., 2017). While the results of this study differ, the findings are supported in one study which reported a Krippendorff’s Alpha value of  $k = 0.38$  (Shultz et al., 2013). This study was unable to conclude if the results were due to the ambiguity of scoring criteria or the need for improved rater training (Shultz et al., 2013). The three clearing tests, being the active straight leg raise, trunk stability push up, and the shoulder mobility tests, have previously demonstrated high reliability (Minick et al., 2010; Onate et al., 2012; Parenteau et al., 2014). As such, the removal of these clearing tests from this study (due to the inability to assess these values via video) may have contributed to the overall lower reliability values found in this study.

Although the FMS™ composite score was not calculated, the FMS™ subtest items displayed considerable variability. Few studies have documented higher variability in the inter-rater reliability of the FMS™ subtest items (Onate et al., 2012; Parenteau et al., 2014; Shultz et al., 2013) with Kappa values ranging as low as 0.26 (Parenteau et al., 2014). A common theme among all studies when examined closely is the lower values for the FMS™ subtest items for the hurdle step, rotary stability and in-line lunge (Minick et al., 2010; Onate et al., 2012; Parenteau et al., 2014; Shultz et al., 2013). Among the 67 raters, during the pretraining assessment, the rotary stability (right and final), overhead squat and inline lunge (right) demonstrated the lowest reliability for the FMS™ subtest items. When re-evaluated after the FMS™ training, the right rotary stability and in-line lunge (final) presented with the lowest reliability scores. These findings are supported by one report which compared the reliability of a small group of novices to experts, and cross-examined the values (Minick et al., 2010). In relation to this study, they reported lower FMS™ sub-item values for the lunge tests and rotary final with values ranging from 0.40 to 0.54 (Minick et al., 2010). In general, the FMS™ subtest items inter-rater reliability values from this study, are lower than previous reports. The subtest items may have a lower value in this report due to the larger sample



size of FMS™ raters or different statistics used. Few studies have performed inter-rater reliability studies with over 20 raters (Gribble et al., 2013; Leeder et al., 2016). In addition, the majority of inter-rater reliability studies have performed intraclass correlation coefficients (ICC) with a 95% confidence interval (Gulgin & Hoogenboom, 2014; Leeder et al., 2016; Onate et al., 2012; Teyhen et al., 2012). The low inter-rater reliability on the FMS™ subtest items is of value because poor reliability of a single item of the FMS™ may negatively influence the FMS™ total score and lead to false assumption (Onate et al., 2012).

Multiple studies have compared and analysed the rater experience level in relation to FMS™ reliability (Gulgin & Hoogenboom, 2014; Minick et al., 2010; Smith et al., 2013; Sorenson, 2016). These studies have not looked at the relationship of experienced versus general experience before and after a FMS™ training intervention. When comparing the inter-rater reliability of uncertified versus certified FMS™ examiners, good reliability has been documented (Onate et al., 2012). One report also notes that certification and FMS™ experience does not seem to improve scoring consistency (Smith et al., 2013). When cross-examining novice versus experienced scorers, Minick et al. (Minick et al., 2010) found that novice scorers demonstrated six excellent levels of agreement and as opposed to expert scorers who achieved only four excellent levels of agreement. To support the discrepancy, another study also revealed that raters with less than 1 year experience had fair inter-rater reliability, and raters with more than 2 year's experience demonstrated poor inter-rater reliability (Shultz et al., 2013). The findings of these reports demonstrate that experienced raters can have lower inter-rater reliability scores than novice examiners which emphasizes the importance of having all FMS™ raters retrained to ensure the highest levels of accurate inter-rater reliability. As summarized (Shultz et al., 2013), when using the FMS™ for clinical purposes to have confidence in the screen, a reliability test should be performed by clinicians and researchers using their own staff and population (Shultz et al., 2013). This study further reiterates the importance of FMS™ training by expert raters in the profession prior to administering and scoring the FMS™ to strengthen the reliability of the screen.

To date, no study has examined the inter-rater reliability after a FMS™ training session. One of the primary aims of this study was to determine whether providing formal training in the FMS™ improved inter-rater reliability. Representative of all examiners in the profession, each examiner in this study had variable levels of training and experience with the FMS™. Contrary to our expectations, when comparing PTI "Raters" following FMS™ training, our findings reported that the inter-rater reliability did not consistently improve, following a FMS™ training course. Although this finding was not significant, the general PTI "Raters" were also compared to a "Master" examiner and the level of agreement between "Rater" and "Master" scores did significantly improve. This finding is particularly important as it demonstrates the inter-rater reliability shift away from a central tendency and towards the correct score. After training, the FMS™ test items increased towards an agreement with the "Master" rater on an average of 22%. As demonstrated in Table 2, after the FMS™ training session, the inter-rater reliability is strengthened towards the correct FMS™ score as provided by a "Master" examiner. Previous inter-rater reliability studies as described in various systematic reviews (Bonazza et al., 2017; Cuchna et al., 2016; Stobierski et al., 2015) have not examined whether raters are able to accurately determine the correct FMS™ score. Practically, this means that if a department were to implement the FMS™ as a predictor of injuries, it should ensure that ample training is provided for assessors. This training could be in the form of FMS™ certification, or through mentorship or supervision by a "master assessor". This will ensure both a high level of agreement between raters for comparisons and greater accuracy of results.

Due to the nature of the study design, which was set up across a two-day annual training session, a limitation of this study may include a learning effect. Participants were exposed to 98 videos twice across the 2 days for pre- and post-assessment and raters may have automatically recalled the videos during the second assessment. Although the videos were randomized, the raters may have recognized the athletes and their movement patterns which could have influenced the FMS™ score provided. Another limitation of this study was the use of two Master raters

who also provided the FMS<sup>TM</sup> education intervention. A final limitation was the use of only four of seven components of the FMS<sup>TM</sup>. Although previous literature has documented moderate to excellent reliability in these areas, due to the removal of three items, the reliability of the FMS<sup>TM</sup> composite score or the other three variables cannot be verified with this study. Video recording to determine inter-rater reliability was considered as a limitation of this study, however upon review of literature, the use of video for FMS<sup>TM</sup> inter-rater reliability has been well documented and has demonstrated excellent reliability (Parenteau et al., 2014; Shultz et al., 2013). Future studies may want to confirm the improved accuracy after an FMS<sup>TM</sup> training intervention, examine reasons for the variable FMS<sup>TM</sup> subtest item inter-rater reliability, and diversify the reliability in other law enforcement and tactical populations.

## 5. Conclusion

The findings of this paper confirm the variability of the FMS<sup>TM</sup> subtest items using a large sample size and emphasize the importance of FMS<sup>TM</sup> training by staff prior to conducting video analysis. It is imperative that clinicians scoring the FMS<sup>TM</sup> are retrained by their own experienced staff members and performing reliability tests against a Master rater prior to evaluating the FMS<sup>TM</sup>. This baseline is essential in ensuring the accuracy of the FMS<sup>TM</sup> scores. Without the confidence of the FMS<sup>TM</sup> scores being accurate, the FMS<sup>TM</sup> cannot be utilized to predict injury or abnormal movement patterns.

### Acknowledgements

The authors have no acknowledgements to make.

### Funding

No funding was obtained for this study.

### Competing interests

The authors have no conflict of interests or competing interests to declare.

### Author details

Charise Conkin<sup>1</sup>  
 E-mail: [cckinpt@gmail.com](mailto:cckinpt@gmail.com)  
 Ben Hinton<sup>2</sup>  
 E-mail: [hint1ben@police.nsw.gov.au](mailto:hint1ben@police.nsw.gov.au)  
 Kayla Ross<sup>2</sup>  
 E-mail: [ross1kay@police.nsw.gov.au](mailto:ross1kay@police.nsw.gov.au)  
 Ben Schram<sup>13</sup>  
 E-mail: [bschram@bond.edu.au](mailto:bschram@bond.edu.au)  
 Rodney Pope<sup>34</sup>  
 E-mail: [rpope@csu.edu.au](mailto:rpope@csu.edu.au)  
 Robin Orr<sup>13</sup>  
 E-mail: [rorr@bond.edu.au](mailto:rorr@bond.edu.au)

<sup>1</sup> Faculty of Health Science & Medicine, Bond University, Gold Coast, QLD 4229, Australia.

<sup>2</sup> Health and Fitness Unit, New South Wales Police Force, Parramatta, NSW 2010, Australia.

<sup>3</sup> Tactical Research Unit, Bond Institute of Health and Sport, Faculty of Health Science & Medicine, Bond University, Gold Coast, QLD, 4229, Australia.

<sup>4</sup> School of Community Health, Charles Sturt University, Albury, NSW 2640, Australia.

### Ethics approval and consent to participate

All participants provided written informed consent, and ethics approval for this study was provided by the Bond University Human Research Ethics Committee under protocol RO 1595. Gatekeeper approvals were provided by the Australian State Police Force from which participants were drawn.

### Availability of data and material

Given the nature of the population, data will not be available without a reach forward requested which will need to be approved by both the host organisation and the Bond University Human Research Ethics Committee.

### Citation information

Cite this article as: Inter-rater reliability and a training effect of the functional movement screen in police physical training instructors, Charise Conkin, Ben Hinton, Kayla Ross, Ben Schram, Rodney Pope & Robin Orr, *Cogent Social Sciences* (2020), 6: 1763769.

### References

- Achterstraat, P. (2008). Managing injured police: NSW police force. *Audit Office of New South Wales*.
- Bock, C., Stierli, M., Hinton, B., & Orr, R. M. (2016). The functional movement screen as a predictor of police recruit occupational task performance. *Journal of Bodywork and Movement Therapies*, 20(2), 310–315. <https://doi.org/10.1016/j.jbmt.2015.11.006>
- Bonazza, N. A., Dhawan, A., Smuin, D., Onks, C. A., & Silvis, M. L. (2017). Reliability, validity, and injury predictive value of the functional movement screen. *American Journal of Sports Medicine*, 45(3), 725–732. <https://doi.org/10.1177/0363546516641937>
- Bonneau, J., & Brown, J. (1995). Physical ability, fitness and police work. *Journal of Clinical Forensic Medicine*, 2(3), 157–164. [https://doi.org/10.1016/1353-1131\(95\)90085-3](https://doi.org/10.1016/1353-1131(95)90085-3)
- Bushman, T. T., Grier, T. L., Canham-Chervak, M., Anderson, M. K., North, W. J., & Jones, B. H. (2015). The functional movement screen and injury risk. *American Journal of Sports Medicine*, 44(2), 297–304. <https://doi.org/10.1177/0363546515614815>
- Butler, R. J., Plisky, P. J., & Kiesel, K. B. (2012). Interrater reliability of videotaped performance on the functional movement screen using the 100-point scoring scale. *Athletic Training & Sports Health Care: The Journal for the Practicing Clinician*, 4(3), 103–109. <https://doi.org/10.3928/19425864-20110715-01>
- Chorba, R. S., Chorba, D. J., Bouillon, L. E., Overmyer, C. A., & Landis, J. A. (2010). Use of a functional movement screening tool to determine injury risk in female collegiate athletes. *North American Journal of Sports Physical Therapy*, 5(2), 47–54. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2953387/>
- Cook, G., Burton, L., & Hoogenboom, B. (2006). Pre-participation screening: The use of fundamental movements as an assessment of function - part 1. *North American Journal of Sports Physical Therapy*, 1



- (2), 62. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2953313/>
- Cuchna, J. W., Hoch, M. C., & Hoch, J. M. (2016). The interrater and intrarater reliability of the functional movement screen: A systematic review with meta-analysis. *Physical Therapy in Sport*, 19, 57–65. <https://doi.org/10.1016/j.ptsp.2015.12.002>
- Decker, A., Orr, R., Pope, R., & Hinton, B. (2016). Physiological demands of law enforcement occupational tasks in Australian police officers. Poster session presented at ASCA National Conference on Applied Strength and Conditioning, Melbourne, Australia.
- Gribble, P. A., Brigle, J., Pietrosimone, B. G., Pfile, K. R., & Webster, K. A. (2013). Intrarater reliability of the functional movement screen. *Journal of Strength & Conditioning Research*, 27(4), 978–981. <https://doi.org/10.1519/JSC.0b013e31825c32a8>
- Gulgin, H., & Hoogenboom, B. (2014). The functional movement screening (FMS)<sup>TM</sup>: An inter-rater reliability study between raters of varied experience. *International Journal of Sports Physical Therapy*, 9(1), 14–20. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3924604/>
- Kiesel, K., Plisky, P., & Butler, R. (2011). Functional movement test scores improve following a standardized off-season intervention program in professional football players. *Scandinavian Journal of Medicine & Science in Sports*, 21(2), 287–292. <https://doi.org/10.1111/j.1600-0838.2009.01038.x>
- Kiesel, K., Plisky, P., & Voight, M. (2007). Can serious injury in professional football be predicted by a preseason functional movement screen? *North American Journal of Sports Physical Therapy*, 2(3), 147–158. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2953296/>
- Leeder, J. E., Horsley, I. G., & Herrington, L. C. (2016). The inter-rater reliability of the functional movement screen within an athletic population using untrained raters. *Journal of Strength & Conditioning Research*, 30(9), 2591–2599. <https://doi.org/10.1519/JSC.0b013e3182a1ff1d>
- Letafatkar, A., Hadadnezhad, M., Shojaedin, S., & Mohamadi, E. (2014). Relationship between functional movement screening score and history of injury. *International Journal of Sports Physical Therapy*, 9(1), 21–27. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3924605/>
- Lisman, P., O'Connor, F., Deuster, P., & Knapik, J. (2013). Functional movement screen and aerobic fitness predict injuries in military training. *Medicine & Science in Sports & Exercise*, 45(4), 636–643. <https://doi.org/10.1249/MSS.0b013e31827a1c4c>
- Lyons, K., Radburn, C., Orr, R., & Pope, R. (2017). A profile of injuries sustained by law enforcement officers: A critical review. *International Journal of Environmental Research and Public Health*, 14(2), 142. <https://doi.org/10.3390/ijerph14020142>
- Minick, K., Kiesel, K., Burton, L., Taylor, A., Plisky, P., & Butler, R. (2010). Interrater reliability of the functional movement screen. *Journal of Strength and Conditioning Research*, 24(2), 479–486. <https://doi.org/10.1519/JSC.0b013e3181c09c04>
- O'Connor, F., Deuster, P., Davis, J., Pappas, C., & Knapik, J. (2011). Functional movement screening: Predicting injuries in officer candidates. *Medicine & Science in Sports & Exercise*, 43(12), 2224–2230. <https://doi.org/10.1249/MSS.0b013e318223522d>
- Onate, J. A., Dewey, T., Kollock, R. O., Thomas, K. S., Van Lunen, B. L., DeMaio, M., & Ringleb, S. I. (2012). Real-time intersession and interrater reliability of the functional movement screen. *Journal of Strength and Conditioning Research*, 26(2), 408–415. <https://doi.org/10.1519/JSC.0b013e318220e6fa>
- Parenteau, G. E., Gaudreault, N., Chambers, S., Boisvert, C., Grenier, A., Gagne, G., & Balg, F. (2014). Functional movement screen test: A reliable screening test for young elite ice hockey players. *Physical Therapy in Sport*, 15(3), 169–175. <https://doi.org/10.1016/j.ptsp.2013.10.001>
- Peate, W. F., Bates, G., Lunda, K., Francis, S., & Bellamy, K. (2007). Core strength: A new model for injury prediction and prevention. *Journal of Occupational Medicine and Toxicology*, 2(1), 3. <https://doi.org/10.1186/1745-6673-2-3>
- Perry, F. T., & Koehle, M. S. (2013). Normative data for the functional movement screen in middle-aged adults. *Journal of Strength & Conditioning Research*, 27(2), 458–462. <https://doi.org/10.1519/JSC.0b013e3182576fa6>
- Schneiders, A. G., Davidsson, A., Horman, E., & Sullivan, S. J. (2011). Functional movement screen normative values in a young, active population. *International Journal of Sports physical Therapy*, 6(2), 75–82. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3109893/>
- Shultz, R., Anderson, S. C., Matheson, G. O., Marcello, B., & Besier, T. (2013). Test-retest and interrater reliability of the functional movement screen. *Journal of Athletic Training*, 48(3), 331–336. <https://doi.org/10.4085/1062-6050-48.2.11>
- Smith, C., Chimera, N., Wright, N., & Warren, M. (2013). Interrater and intrarater reliability of the functional movement screen. *Journal of Strength and Conditioning Research*, 27(4), 982–987. <https://doi.org/10.1519/JSC.0b013e3182606df2>
- Sorenson, E. A. (2016). The reliability of functional movement screen scores assigned by athletic trainers. *International Journal of Athletic Therapy & Training*, 21(3), 38–41. <https://doi.org/10.1123/ijatt.2015-0070>
- Stobierski, L. M., Fayson, S. D., Minthorn, L. M., McLeod, T. C., & Welch, C. E. (2015). Reliability of clinician scoring of the functional movement screen to assess movement patterns. *Journal of Sport Rehabilitation*, 24(2), 219–222. <https://doi.org/10.1123/jsr.2013-0139>
- Teyhen, D. S., Shaffer, S. W., Lorenson, C. L., Halfpap, J. P., Donofry, D. F., Walker, M. J., Dugan, J. L., & Childs, J. D. (2012). The functional movement screen: A reliability study. *Journal of Orthopaedic & Sports Physical Therapy*, 42(6), 530–540. <https://doi.org/10.2519/jospt.2012.3838>
- Wright, M. D., Portas, M. D., Evans, V. J., & Weston, M. (2015). The effectiveness of 4 weeks of fundamental movement training on functional movement screen and physiological performance in physically active children. *Journal of Strength & Conditioning Research*, 29(1), 254–261. <https://doi.org/10.1519/JSC.0000000000000602>



© 2020 The Author(s). This open access article is distributed under a Creative Commons Attribution (CC-BY) 4.0 license.

You are free to:

Share — copy and redistribute the material in any medium or format.

Adapt — remix, transform, and build upon the material for any purpose, even commercially.

The licensor cannot revoke these freedoms as long as you follow the license terms.

Under the following terms:

Attribution — You must give appropriate credit, provide a link to the license, and indicate if changes were made.

You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.

No additional restrictions

You may not apply legal terms or technological measures that legally restrict others from doing anything the license permits.

**Cogent Social Sciences (ISSN: 2331-1886) is published by Cogent OA, part of Taylor & Francis Group.**

**Publishing with Cogent OA ensures:**

- Immediate, universal access to your article on publication
- High visibility and discoverability via the Cogent OA website as well as Taylor & Francis Online
- Download and citation statistics for your article
- Rapid online publication
- Input from, and dialog with, expert editors and editorial boards
- Retention of full copyright of your article
- Guaranteed legacy preservation of your article
- Discounts and waivers for authors in developing regions

**Submit your manuscript to a Cogent OA journal at [www.CogentOA.com](http://www.CogentOA.com)**

